Does embodied vision unravel at the movies?

By Lenhardt Stevens

Introduction

Embodied cognition as an approach garners its appeal by suggesting that an agent's interactions with its environment are inextricably linked to how we understand its cognitive processes; you cannot take them apart without somehow distorting our understanding. Thus, the agent, coupled with their environment, becomes a dynamically interconnected system.[1] Enactivism takes it one step further. I intend to pose a series of problems to the enactivist camp by sketching an answer to the question: how can we have an embodied approach to watching movies? In addressing this issue, I will be alongside Jesse Prinz[2] and Andy Clark[3] in an effort to assemble some previously explicated shortcomings with the enactivist camp's contentions (I am referring primarily to Hutto & Myin[4]; Alva Noë[5]). Broadly construed, the problem will be with the notion that our perceptual experience of a movie could ever fall under the doctrine of perception purely for action.[6] In the interest of space, I will not pin down appropriate solutions to these problems outright, but rather refer the reader to positions that will give us a better account of how visual experience at the movies can be made sense of in light of these issues within embodied cognition.

Commented [AM1]: Although, experiences of interaction with the world are usually plentiful before we perceptually experience our first movie…whether they are essential or not may be the issue

[1] Early promises in this theoretical domain are plentiful, but I am referring mostly to Thelen, E., and L.B., Smith, 1994, *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge, MA: MIT Press.

[2] Prinz, Jesse (2006) "Putting the Brakes on Enactive Perception." *Psyche 12*. March.

[3] Clark, Andz. (2001) "Visual Experience and Motor Action: Are The Bonds Too Tight?" *Philosophical Review* 110:4 October.

[4] Hutto & Myin. (2013) *Radicalizing Enactivism: Basic Minds without Content*. MIT Press.

[5] Noë, A., 2004, *Action in Perception*, Cambridge, MA: MIT Press

[6] Hutto's definition (2010, p. 3): "Enactivist approaches treat mentality as an emergent phenomenon that is constituted by, and thus to be understood in terms of, specifiable developing patterns of interaction between organisms (natural or artificial) and aspects of their environments."

Theater-Seated, Movie-thinking

What perceptual systems are running in the viewer while the movie plays? Rewind to a world of silent films, when all the information from the movie is on the screen. I make four assertions that divide the labor for what the embodied cognitivist must address in order to give a proper account of vision at the movies. They are:

i) We cannot *physically interact* with the films represented environment, characters, or events, yet we *comprehend* their presentation because of prior knowledge of certain *cinematically characteristic spatio-temporal relationships and their respective continuities*, thereto

ii) there are ways in which the visual experience we have at the movies *resembles* the visual experiences we have with the everyday world.

iii) The viewer *knows* that they cannot interact with the moving images on screen, and that they do not have any bearing on immediate circumstances of the real world.

therefore

iv) visual representation, in some important sense, is still required in order to explain visual experience obtained while watching a movie

Visual representation here means the mental processes that enable us to perceive visual stimuli and are not constitutive of the mental processes that enable us to act, a complete departure from enactivism.[7] Although the premises are interrelated, they can bear independence from one another. i) and ii) attempt to describe the *similarities* in visual experience between movies and real-life, whereas iii) is an epistemic position concerned with the way in which we experience films as separate from our immediate circumstances, e.g. the swinging battle axe wielded by Achilles cannot cut me while I am in my theater chair. Endorsing ii) entails i), but I could maintain i) without holding ii). How? Say I am a being who lives a life interacting with objects in three-dimensional space and an understanding a linear progression of causes, but I also observe things called "super-movies," which cannot be understood using any of my prior knowledge of

[7]Prinz (2006) p. 3

Commented [AM2]: Not sure we have to understand any cinematic conventions or practices to see a movie..rats can be shown movies of natural scenes and sem abel to understand what they are seeing…

space and time in my normally action-driven environment. Instead, understanding these super-movies requires a totally different set of cognitive mechanisms, which make me react and register these visual phenomena in distinct ways from how I interact with my normal environment. For that reason, I will be discussing primarily premise (i) in this essay, knowing that premise (ii) will fall out of it if argued correctly. Even if you wanted to assert that movies are analogues to dreams, in that we misapprehend their reality, thereby rejecting premise (iii), you would still need to address how that illusion is perceptually possible.[8] The point is this: our understanding of movies flatly denies an enactivist position as being relevant for our experience of them, because none of the visual information has the capacity to motivate action on the part of the spectator towards the film. Animate Vision,[9] viz. of the sort endorsed by enactivism, does not account for our experience at the movies.

Commented [AM3]: I'm not sure this is really imagineable. Even video games that warp the laws of physics bear a resemblance to the real world

Commented [AM4]: But watching a movie involving dance, say, will activate many of the neural areas that are involved in motor control?

Fluent motor interaction and physical presence are both impossible with cinematic space, yet we retain the here-and-now apprehension of the visual information. Disengaging from movie events in such a way as to not fear for my personal safety while Achilles' swings his battle-axe, yet still being able to speak of bearing witness to an extemporally present axe swinging Achilles is the crux of the issue.

Continued Shortcomings of Radical Enactivism

The images captured by the film camera retain the salient features of a visual scene that humans need to make sense of it.[10] The play of light and their relevant sources make shadows and the brightness of colors captured by the film camera appear in such a way that our eyes have sufficient information to make sense of a three-dimensional world from a two-dimensional succession of images. Biedermann's Geon theory, among others, demonstrates that identifying vertices in a two-dimensional image grants the

[8] Dreams for the radical enactivist are another major counter example for the project; see Prinz (2006), p. 17
[9] Clark (1999), p. 4
[10] Anderson, Joseph A. (1998) *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory.* Southern Illinois University Press.

viewer the knowledge of volumetric geometry, hence our easy inferences about three-dimensional objects on a two-dimensional plane.[11]

But the issue arises over whether or not we have completely or only partially disengaged our vision-for-action. That is, there is some kind of selection at work in our neural architecture so that information about the couch we are reclined upon stays where it is whereas visual information from the television is not used for these spatial considerations. If this last statement were false, then we would have to suggest that our perceptual experience feels as if it could reach into the television and start picking things out in the visual image.[12]

On-line object-engaged performance information is simply a superfluous postulate at the movies, and it would be a waste of cognitive resources to attempt engaging it. Our conscious experience, comprehending multiple and even inconsistent visual components, i.e. the television is flat yet its images are of people in realizably three-dimensional space, helps us understand the reason for this comprehension.[13] Finely tuned details about object size might be irrelevant, but it means that things like retinal image size is something important for conscious visual experience insofar as we are able to use it to determine object relationships in a scene's visuo-spatial array. Think about the fact that your watching *Godzilla* on your smartphone, tablet, PC, home theater, or iMax screen simply does not affect the inferences you are able to correctly make about the movie; the ecological context *qua* the space in which I watch the movie is irrelevant for cinematic information. This does not mean, however, that J.J. Gibson's ideas on ecology for perception get tossed under the rug. On the contrary, for we have reason to agree, *a la* premise (ii), that "(t)he progressive picture [i.e., movies] yields something closer to natural visual perception than does the arrested image [i.e., photographs]."[14] Super-movies be damned.

---

[11] Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.
[12] See Soreno (2006) for experimenting with television images and neural activation.
[13] Clark (2011), p. 20

[14] Gibson, J.J., 1979, *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin, p. 302

Films as a slice of Reality

We now get back to the general topic of the essay. Can we still talk about embodied cognition as something that gets us an account of visual perception of movies? Is the vision we may associate with action-oriented mental processing responsible for any of this experience that we are having? I think so.

Shared facts between the world as an interaction-possible space and the projected movie are bountiful.[15] It is certainly true of both the world and movies that their persisting outside of our attending allows us to offload computation to the film. Watching *Chinatown* (with the sound off), you can recognize that on the day the scene at the reservoir was shot Jack Nicholson was wearing a suit, the sun was shining, and they are somewhere outside of Los Angeles. You obtain all of this information from visual stimuli that contains identifiable conceptual features shared with a sunny day spent on a set with Jack Nicholson outside Los Angeles. In line with Clark, you would not have had to infer from any jarringly unique pieces of evidence in order to come to this conclusion, rather, your previous model of the world could be used to understand the context of the cinema-seat *and* the information from the movie-screen. The percepts engaged in order to activate real-world behaviors are disengaged, and a simulation is running while applying appropriate, context-sensitive inferential rules that already govern my understanding of the world I inhabit. In other words, we have two models running; one of the film world and another of the world in which we are viewing the film. The environment within the film changes and we adjust our comprehension of events according to it.

We would have much of the information we need to meaningfully interact with the environment depicted within the film, and yet we do not have the same *kind* of information derived from actually *being in* the world of the film. Something has been depreciated as a result of the images on screen being our only way to understand the objects in the environment on screen. We still need the agent embedded within an environment to understand movies, but we also need them to possess a model that can handle representational content devoid of action-oriented concepts.

---

[15] Anderson (1998), p. 25

Interestingly, there is room for visual trickery at the movies that could appeal to illusion-based receptivity. 3D movies appear to be cashing out online information in a way the two-dimensional images cannot. The incoming knife thrown by the ninja assassin might make some people jump in the two-dimensional screening, but only the three-dimensional version makes some audience members move their heads out of the way of the incoming blade. I look forward to exploring this behavioral component of 3D films in the future, as might it be a case of exploiting our reflexive vision-for-action systems.[16]

There will certainly be more desiderata for this account that I must forgo here. For instance, how is it that we have one model used in various contexts if I can have concurrently running predictions of the movie-theater context *and* the drama unfolding onscreen? Information from one impacts the real world, and information from another impacts only the film world. Perhaps we do not want to draw too strict a dividing-line here. Nevertheless, the differences in context relevant stimuli are striking and needs an explanation.

Conclusion

In this paper, I have made a case for movies acting as a counterexample for enactivist commitments to visual perception. Our understanding of them as such puts a hitch in the conversation of radically embodied vision. Hardline enactivists just do not consider the multifarious contexts human vision participates within, and if they did, they would discover their tenets create hurdles only the stubbornly committed could ignore. This is not the same thing as a defeat for embodied vision. The delineation of our senses role for perception and action continues to require representations of the world, in no way suggesting something like "pure vision" wins out.[17]

---

[16] Haggard, P., 2005, "Conscious intention and motor cognition," *Trends in Cognitive Science*, 9 (6) p. 293

[17]Churchland, Ramachandran & Sejnowski (1994) "A Critique of Pure Vision" in Koch, C. and Davis, J. (eds.) *Large-Scale Neuronal Theories of the Brain*, pp. 23-60 Cambridge, MA: MIT Press, 1994.

Future applications of embodied vision will have to take this under consideration if the aim is for the senses to resemble a human visual system. Robotics will have to develop visual systems that can look at television screens and understand the progressive pictures displayed on them in a way that makes sense to achieve an accurate visual representation. The agent must be able to watch *The Simpsons* and know that it does not possess the possibility of sharing a beer with Homer while Homer is having a beer. If we spend too much time embodying visual systems in robots, we may miss this crucial aspect of human visual perception. It seems traditional visual systems have an edge over the embodied systems in this respect, as the cues in classical systems, e.g. objects classified under programmable and symbolically interpretable registers, should work, at least minimally, for how cinematic images and their events are understood by human spectators.[18]

Commented [AM5]: I worry that you make too much of this. My ability to perceive the events on screen could be inextricably bound up with my abilities to act in the world, even if I also learn that I can't – for other reasons – interact with the depicted events. There are discussions in Noe of how enavtivism is consistant with the fact that paralysis, despite blocking interactions, need not cause blindness...you could see if these replies apply to the movie case too.

We should continue conducting trials to understand how humans who watch movies make inferences about spatial relations between objects within the video footage. If we are able to make sense of the basketball game on TV, we can do so in virtue of cortical activation that does not constitute any sensorimotor capacities. Patients who have optic ataxia, I speculate, can probably enjoy the game or movies in a way that sufferers of other visual agnosia cannot.[19]

Finally, vision is only one ingredient in the dish of sensory experience. Auditory information, and its role in locating some object within cinematic space, will respond to many of these same issues. Further developments in theories of perception at the cinema should account for them, as well.

---

[18] Marr, D., 1982, *Vision: A Computational View*, San Francisco: Freeman Press.

[19] McIntosh, RD.; Schenk, T. (May 2009). "Two visual streams for perception and action: current trends.". *Neuropsychologia* 47 (6): 1391–6.

Works cited

Sereno MI & Huang RS. (2006) ”A human parietal face area contains aligned head-centered visual and tactile maps.” Natural Neuroscience. Oct;9(10):1337-43.

Singh-Curry, V. & Husain, Masud.”The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy”.

Churchland, Ramachandran & Sejnowski (1994) “A Critique of Pure Vision” in Koch, C. and Davis, J. (eds.) *Large-Scale Neuronal Theories of the Brain*, pp. 23-60 Cambridge, MA: MIT Press, 1994.

Donald, M., 1991, *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*, Cambridge, MA: Harvard University Press.

Gibson, J.J., 1979, *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin.

Haggard, P., 2005, “Conscious intention and motor cognition,” *Trends in Cognitive Science*, 9 (6): 290–295.

Marr, D., 1982, *Vision: A Computational View*, San Francisco: Freeman Press.

McIntosh, RD.; Schenk, T. (May 2009). "Two visual streams for perception and action: current trends.". *Neuropsychologia* 47 (6): 1391–6.

Anderson, Joseph A. (1998) *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Southern Illinois University Press.

Thelen, E., and L.B., Smith (1994) *A dynamic systems approach to the development of cognition and action*, Cambridge, MA: MIT Press.

Prinz, Jesse (2006) “Putting the Brakes on Enactive Perception.” *Psyche 12*. March.
Clark, Andz. (2001) “Visual Experience and Motor Action: Are The Bonds Too Tight?” *Philosophical Review* 110:4 October.
Hutto & Myin. (2013) *Radicalizing Enactivism: Basic Minds without Content*. MIT Press.
Noë, A., 2004, *Action in Perception*, Cambridge, MA: MIT Press

Goodale MA, Milner AD (1992). "Separate visual pathways for perception and action". *Trends Neurosci.* 15 (1): 20–5

Hutto, D. (2010) “Enactivism: Why Be Radical?” Published online at: http://www.academia.edu/245547/Enactivism_Why_be_Radical

Lorch, E. P., Anderson, D. R., & Levin, S. R. (1979) The relationship of visual attention to children's
comprehension of television. *Child Development*, 50, pp. 722-727.

Clark, Andy (1999). ""From Fish to Fantasy: Reflections on an Embodied Cognitive Science" A shortened and amended version appears as "An Embodied Cognitive Science?" *Trends In Cognitive Sciences* 3:9:1999 p. 345-351

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.